Why I am not a physicalist

Michael Pelczar

We consider four types of broadly anti-physicalist argument: arguments purporting to show that physicalism is false, arguments purporting to show that we should believe that physicalism is false, arguments purporting to show that we should not believe that physicalism is true, and arguments purporting to show that we do not, in fact, believe that physicalism is true. For the physicalist, there's good news and bad news, but mostly bad. The good news for the physicalist is that arguments of the first type pose no clear and present danger to his position. The bad news is that each of the other three types of argument has at least one instance that the physicalist has no evident way of blocking.

1 Introduction

This paper considers a variety of arguments that aim to put pressure on physicalism. Some of the arguments purport to show that physicalism is false, some that we should believe that physicalism is false, some that we should refrain from believing that physicalism is true, and some that we do, as a plain matter of fact, believe that physicalism is false (or, fail to believe that it's true). The paper proceeds as follows.[1]

[1]I understand physicalism as the view that the physical facts metaphysically entail all the facts. There are various ways of formulating physicalism more precisely, but

§2 reviews arguments purporting to show that physicalism is false, with a focus on so-called modal or conceivability arguments. The conclusion of this section is that those arguments are not very successful, at least from a polemical standpoint.

§3 considers an argument for the claim that we should believe that physicalism is false. The argument, though promising, is flawed.

§4 modifies the flawed argument of §3 so that it concludes—not that we should believe that physicalism is false, but—that we shouldn't believe that physicalism is true. I contend that it's extremely difficult for a physicalist to resist this argument in good faith.

§5 considers a different modification of the argument of §3. This one escapes the objection to the flawed argument, while still concluding that we should believe that physicalism is false. I contend that this argument is also very hard to resist in good faith.

§6 considers an argument that concludes that we do not, in fact, believe that physicalism is true. This is an amended version of an argument that David Papineau attributes to Saul Kripke. I agree with Papineau that the Kripkean argument is sound, but I argue that Papineau underestimates the challenge it poses to would-be physicalists.

---

the purpose of those refinements is to overcome technical difficulties orthogonal to the arguments of this paper; see (Chalmers, 1996, 38-41), (Jackson, 1998, 9-14), (Stoljar, 2010, 133-39), and (Blumson & Tang, 2015).

2 Arguments against physicalism

There are two main ways to try to show that physicalism is false: via the knowledge argument, and via modal arguments. Here, my focus will be on modal arguments.[2]

Modal arguments attempt to deduce the falsity of physicalism from two premises: one to the effect that we can (under certain specified circumstances) conceive of a world that duplicates ours physically, but fails to contain all of the conscious experience that ours contains, and another to the effect that if we can conceive of such a world (under the specified circumstances), such a world is metaphysically possible. The simplest modal argument goes like this:

A1 We can conceive of a world that duplicates our world in all physical respects, but contains no conscious experience.

A2 If we can conceive of such a world, then such a world is metaphysically possible.

A3 So, it's metaphysically possible for there to be a physical duplicate of our world that contains no conscious experience.

Let's call this *the naive modal argument* or "naive argument," for short. Most people (including me) are willing to grant its first premise. We can conceive of a world without rubber that contains all the metallic stuff that our world contains, arranged just as it is in our world; it seems no

[2]Elsewhere, I've argued that we have good reasons to suspend judgement on whether the knowledge argument is sound. I'm not going to discuss those reasons here.

harder to conceive of a world without consciousness that contains all the physical stuff our world contains, arranged just as it is in our world.

However, as many have pointed out, the naive argument's second premise is suspect. The problem is that people sometimes conceive of things that are, it turns out, metaphysically impossible. This happens whenever someone takes himself to have proved a mathematical proposition that turns out to be false. For example, Hobbes notoriously claimed to have solved the problem of squaring the circle; presumably, he conceived of himself as having squared the circle. But such an achievement is demonstrably impossible. Arguably, something similar happens whenever someone misidentifies a natural kind. At one time it was an open question whether water was H_2O or just HO. 18th century chemists in the HO camp presumably conceived of water as being HO. But given (as is nowadays generally accepted) that it's metaphysically necessary that water is H_2O, those chemists conceived of something metaphysically impossible.[3,4]

The standard response to this criticism is to replace the naive modal argument with a more sophisticated argument, which I'll call *the refined modal argument*, or "refined argument" for short:[5]

---

[3]For other cases of this sort, including the time-travel case discussed below, see (Kung, 2010).

[4]Here and throughout, by "it's metaphysically necessary that water is H_2O," I mean that it's metaphysically necessary that *if water exists,* then water is H_2O; obviously, it's not the case that water is H_2O in worlds in which water doesn't exist.

[5]See (Chalmers, 1996, 93-171) and (Chalmers & Jackson, 2001). What follows is just one example of a refined modal argument against physicalism. There are others; for example, there's the inverted-spectrum argument that replaces "world that contained no conscious experience" with "world in which people's visual experience

B1 We could conceive of a physical duplicate of our world that contained no conscious experience, even if we knew all the logical, mathematical, and microphysical facts.

B2 If we could conceive of a physical duplicate of our world that contained no conscious experience even if we knew all the logical, mathematical, and microphysical facts, then such a world is metaphysically possible.

B3 So, it's metaphysically possible for there to be a world that duplicates our world in all physical respects, but contains no conscious experience.

Maybe some 18th century chemists conceived of water as being HO, but they wouldn't have been able to do so if they had known all the microphysical facts, since in that case they'd have known that water consisted of H_2O molecules, and if they'd known that, they'd have no more been able to conceive of water as being something besides H_2O than to conceive of water being something besides water. (Maybe some of the chemists would have believed that they could still conceive of a world in

---

was color-inverted relative to their actual experience." Such a world would fail to contain all of the experience that actually exists, since, for example, it would fail to contain the phenomenally blue experience I have when I look at the sky, containing phenomenally yellow experience instead. A more significant departure from the naive modal argument that still falls within the broad category of conceivability arguments against physicalism is Philip Goff's transparency argument (Goff, 2017, 106-25). The remarks I'm about to make on the refined modal argument also apply to the transparency argument: it is polemically weak, since a physicalist is apt to insist that in our present state of ignorance about the physical, we're not in a position to say that there is an unbridgeable epistemic gap between the experiential and the purely physical.

which water was HO, mistaking their conception of a world in which stuff that outwardly resembles water is HO for a conception of a world in which water is HO, but that just shows that it's possible to think you can conceive of something that you can't, in fact, conceive of.) Likewise, if Hobbes conceived of himself as having squared the circle, that's only because he didn't know all the logical and mathematical facts; if he had known all the logical and mathematical facts, he'd have known that the statement that he squared the circle entailed a contradiction.

Personally, I think the refined argument is sound. But it's polemically weak, since a physicalist is apt to consider its first premise question-begging. How do we know that in our present state of ignorance, we aren't like the chemists who conceived of water as HO, or the philosopher who thought he had squared the circle? True, we can't see how adding to our existing store of physical information could ever prevent us from conceiving of a zombie world, but neither could pre-Cantorian mathematicians see how adding to their store of mathematical informatoin could ever prevent them from conceiving of a highest cardinality, and we know how that story goes.[6,7]

---

[6]The point is a familiar one; see, e.g., (Block & Stalnaker, 1999, 13-16).

[7]A physicalist might also try to cast doubt on the refined argument by pointing out that an ancient astronomer who thought that a transcendent immaterial deity was responsible for the the movements of celestial bodies might have used the same style of argument to "prove" that a physical duplicate of our world could be one in which nothing explained planetary motion. The astronomer would reason that even if he knew all the logical, mathematical, and microphysical facts, he could conceive of a physical duplicate of our world that contained no such deity, and consequently no explanation of planetary motion; from this, he would infer (via the principle that conceivability in the light of all logical, mathematical, and microphysical information ensures metaphysical possibility) that planetary motion could be inexplicable

There is one suggestive disanalogy between the chemical and mathematical cases and the case of consciousness, familiar from Kripke.[8]

Before the HO vs. H_2O controversy was settled, chemists, like everyone else, had a way of distinguishing water from non-water (grain alcohol, nitric acid, etc). There are certain features such that prior to settling the controversy, chemists classified samples as water or non-water accordingly as the samples had or lacked those features. Let's call these the "outward features" of water (although they needn't be evident to the casual or untrained observer), and let's say that anything that has these features "outwardly resembles" water.

An astute 18th century chemist in the water-is-HO camp might have reflected that if, contrary to his hypothesis, water turned out to be H_2O, then what he was tempted to describe as "conceiving of water as HO" might really just be conceiving of a situation in which something that outwardly resembles water is HO, but not a situation in which water is HO. Then he'd have backed off from his claim to conceive of water as HO, and suspended judgement on the question of what, exactly, he was conceiving of.

---

in a world physically indistinguishable from ours. Of course, such a world is not possible, as we've known since Kepler. The anti-physicalist has an easy response to this, however: unlike the ancient astronomer's deity, conscious experience is not just an explanatory posit, but a datum of (inner) observation. The most that the case of the ancient astronomer shows is that we should apply the relevant principle (that conceivability in the light of all logical, mathematical, and microphysical information entails metaphysical possibility) only in cases in which what's at issue is something that's not just an explanatory posit.

[8]See (Kripke, 1980, 144-54).

An astute 21st century philosopher in the consciousness-is-brain-activity camp can't engage in parallel reflections. He can't reflect that if his hypothesis is wrong and consciousness turns out to be something non-physical, then what he's tempted to describe as "conceiving of consciousness as brain-activity" might really just be conceiving of a situation in which something that outwardly resembles consciousness is brain-activity, but not a situation in which consciousness is brain-activity. If he's conceiving of a situation in which something that outwardly resembles consciousness is a brain-activity, then he *is* conceiving of a situation in which consciousness is brain-activity, since anything that outwardly resembles consciousness is consciousness.

After all, for something to bear an outward resemblance to consciousness is just for it to have whatever features prompt us to identify certain things as instances of consciousness, prior to the resolution of the debate over whether consciousness is brain-activity. But those features are phenomenal features—features by virtue of having which a state is such that there's something it's like to be in it (or a process such that there's something it's like to undergo it). It's true that my evidence that an instance of consciousness is present often consists of another person's behavior, but what I take the behavior to be evidence *of* is something that the behaver can identify as an instance of consciousness in the same way that I identify instances of consciousness in me, namely by their possession of phenomenal features.

Suggestive as it is, it's not obvious how the disanalogy between these cases (water versus consciousness) helps proponents of the refined modal argument avoid the charge of question-begging. A physicalist can grant

that anything that outwardly resembles consciousness *is* consciousness, but insist that for all we know, there are complex neural processes that outwardly resemble consciousness, so that detailed knowledge of those processes would prevent us from conceiving of them as occurring in the absence of conscious experience. We can't presently conceive of such processes, but there was a time when people couldn't conceive of sets with too many members to fit onto an endless list. Maybe, when it comes to consciousness, we are like those people.[9,10]

3 An argument for believing that physicalism is false

Earlier, we saw that the naive modal argument failed, due to known cases in which someone conceived of something that turned out to be impossible. Still, the claim that conceivability lends credence to metaphysical

[9]Or like Daniel Stoljar's intelligent slugs, who inhabit a mosaic comprising triangular and slice-of-pie-shaped tiles combined in various geometric patterns, including circular patterns (Stoljar, 2006, 3-13). Due to limitations on their perceptual apparatus, the slugs can detect only triangles and circles. Dualist slugs contend that circular features of the mosaic don't supervene on the arrangement of its fundamental geometric constituents, on the grounds that they can conceive of a mosaic that duplicates theirs in its micro-tesselar properties, but contains no circles. The dualist slugs are wrong, of course, and if they knew why—viz., because their mosaic includes certain non-triangular tiles that they're unable to detect—they'd lose their ability to conceive of a mosaic that duplicated theirs without containing any circles. Analogously, Stoljar suggests, our ability to conceive of zombies may be due to our being "unaware of a type of nonexperiential truth relevant to the nature of experience." (Stoljar, 2006, 87ff) To the objection that, unlike the slugs, we can't even *imagine* acquiring physical information that would enable us to construe consciousness as a purely physical phenomenon, it may be replied that prior to Gauss, people couldn't imagine acquiring physical information that would enable them to construe space as curved.

[10]In §6, we'll consider Kripkean arguments that put pressure on physicalists by means of the disanalogy between water/H_2O and pain/C-fiber stimulation.

possibility had some initial plausibility, and it's worth considering where this came from.

I suggest that the claim's initial plausibility comes from everyday reasoning about what's possible. We all believe that there are many ways the world could have been different from how it actually is. The gravitational constant needn't have had exactly the value it actually has, life might never have evolved, a Democrat could have won the presidential election of 2016. Why do we think these things are metaphysically possible? Not because we can point to actual confirming instances of them. It looks as though we believe these things are metaphysically possible because we can conceive of them, *and know of no proof that they are metaphysically impossible.*

Whenever we have a good reason to accept something as an example of someone who conceives of a metaphysically impossible state of affairs (or putative state of affairs), the example is a case of someone who conceives of a putative state of affairs that we know to be provably impossible (e.g., describable in a way that reveals a hidden incoherence or contradiction). For instance, since we can derive a contradiction from the proposition that Hobbes squared the circle, or at least have it on good authority that a contradiction is derivable from that proposition, we don't think that Hobbes' ability to conceive of himself as squaring the circle supports the claim that it's metaphysically possible to square the circle. But if we didn't know that squaring the circle was demonstrably impossible, we couldn't point to Hobbes as a counterexample to the claim that conceivability implies metaphysical possibility. Similarly, if you think that our ability to conceive of time travel shows that conceivability doesn't imply

metaphysical possibility, that's only because you think that time travel is demonstrably impossible. If you didn't think you knew of any way to prove that it was metaphysically impossible for Doc Brown to travel from 1985 to 1885 (e.g., by deriving a contradiction from the claim that such a voyage took place), you wouldn't take our ability to conceive of Doc Brown travelling from 1985 to 1885 as a counterexample to the claim that conceivability implies metaphysical possibility.

The suggestion here is that there's a reasonable presumption that the conceivability of a putative state of affairs establishes its metaphysical possibility, absent any knowledge of a proof that that state of affairs is metaphysically impossible. That's why we consider ourselves justified in believing that the laws of physics or the initial conditions of our universe could have been different from what they actually are: we can conceive of the laws having been different or (for example) matter having been differently distributed circa the Big Bang, and we know of no proof that the laws or distribution couldn't possibly have been different.

If this is right, we have to choose between (1) admitting that conceivability undefeated by knowledge of a proof of impossibility is enough to justify a belief in metaphysical possibility, and (2) suspending judgement on the question of whether it's metaphysically possible for the distribution of matter circa the Big Bang to have been different from what it actually was (or for the gravitational constant to have had a value slightly different from its actual value, or whatever).

Personally, I don't suspend judgement on these questions. If you're tempted to claim that you do, please reflect that we don't generally wonder why it's not the case that p, unless we believe that it's metaphysically

possible that not-p. I don't wonder why seventeen isn't greater than a million, because I don't believe that seventeen could possibly have been greater than a million. So, if you do wonder why our world doesn't have different laws from the actual ones, or why matter wasn't distributed in a less improbable way around the time of the Big Bang, you do think that the laws or distribution could have been different, at least as a matter of metaphysical possibility.

Let's call conceivability in the absence of knowledge of any proof that what's conceived of is metaphysically impossible "undefeated conceivability." The suggestion we're entertaining is that undefeated conceivability justifies belief in metaphysical possibility. If the suggestion is right, we can fix the naive modal argument by slightly weakening its conclusion, as follows:[11,12]

C1 We can form an undefeated conception of a world that duplicates our world in all physical respects but contains no conscious experience.

C2 If we can form an undefeated conception of such a world, we should believe that such a world is metaphysically possible.

[11]Undefeated conceivability is a form of what Chalmers calls "prima facie conceivability" (Chalmers, 2002, 147).

[12]Strictly speaking, the conclusion of this argument isn't weaker than the conclusion of the naive argument, since "a zombie world is metaphysically possible" doesn't logically entail "we should believe that a zombie world is metaphysically possible." But the latter claim is weaker than the former in the sense that the conjunction of "a zombie world is metaphysically possible" and "we should believe what is true" entails "we should believe that a zombie world is possible," whereas the conjunction of "we should believe that a zombie world is possible" and "we should believe what is true" doesn't entail "a zombie world is possible."

C3 So, we should believe that it is metaphysically possible for there to be a physical duplicate of our world that contains no conscious experience.

Call this the *naive untenability argument*. Although it has some initial plausibility, it is subject to the following objection.

According to the Goldbach conjecture, every even number greater than 2 is a sum of two prime numbers. So far, no one has proved or disproved this conjecture. Consequently, I can conceive of the Goldbach conjecture being true; that is, I can imagine it turning out that every even number greater than 2 is a sum of two primes. Furthermore, I know of no contradiction or incoherence in the claim that the Goldbach conjecture is true; if I did, I'd probably be in line for a Fields Medal. According to the principle that undefeated conceivability justifies belief in metaphysical possibility, it follows that I should believe that it's metaphysically possible that the Goldbach conjecture is true.

However, I can also conceive of the Goldbach conjecture being false; that is, I can imagine it turning out that *not* every even number greater than 2 is a sum of two primes. I know of no contradiction or incoherence in the claim that the Goldbach conjecture is false; if I did, I'd be in a position to prove the conjecture (or direct you to a proof of it). According to the principle that undefeated conceivability justifies belief in metaphysical possibility, it follows that I should believe that it's metaphysically possible that the Goldbach conjecture is false.

The upshot is that the principle that undefeated conceivability justifies belief in metaphysical possibility implies that I should believe both that

it's metaphysically possible that the Goldbach conjecture is true, and that it's metaphysically possible that the Goldbach conjecture is false. But I know a priori that it's not both metaphysically possible that the conjecture is true and metaphysically possible that the conjecture is false. As a mathematical proposition, the conjecture is either necessarily true or necessarily false—true in all metaphysically possible worlds or none.[13]

At this juncture, there are two paths forward for the anti-physicalist. One leads to an argument that we should refrain from believing that physicalism is true, the other to an improved argument that we should believe that physicalism is false. Let's start with the former argument.

4 Why we shouldn't believe that physicalism is true

Even though the Goldbach conjecture and similar examples show that it's false that undefeated conceivability justifies belief in metaphysical possibility, they don't threaten a weaker version of the principle. They don't cast doubt on the claim that if you can conceive of its being the case that p, and you don't know of any proof that it's metaphysically impossible that p, then you should *not* believe that p is metaphysically *impossible*. Call this the *suspension principle*.

The suspension principle implies that I shouldn't believe that it's metaphysically impossible that the Goldbach conjecture is true; that's because I can conceive of it being true, and I know of no proof that it's metaphysically impossible for it to be true. For parallel reasons, the prin-

[13] The use of a mathematical example here is inessential. We could have made the same point with an a posteriori necessity, such as the identification of water and H_2O, or Phosphorus and Hesperus.

ciple says that I shouldn't believe that it's metaphysically impossible that the Goldbach conjecture is false. In both cases, the principle gets it right: I shouldn't believe either that there are no possible worlds in which the conjecture is true or that there are no possible worlds in which the conjecture is false. I should suspend judgement on the conjecture's possible truth or possible falsity, which is just what the suspension principle recommends.

This gives us the following argument against believing physicalism:

D1 We can form an undefeated conception of a zombie world (a world that duplicates ours in all physical respects, but contains no conscious experience).

D2 If we can form an undefeated conception of such a world, we shouldn't believe that such a world is metaphysically impossible.

D3 So, we shouldn't believe that it's metaphysically impossible for there to be a zombie world.

Call this the *suspension argument*. What it sacrifices in the strength of its conclusion, it gains in irresistibility.

We think that things never exceed the speed of light in vacuo (c), but we don't think that it's metaphysically necessary that this is so. Why not? Because we can conceive of something moving faster than c, and know of no proof that superluminal travel is metaphysically impossible. (Any reason for having the positive belief that superluminal travel *is* metaphysically possible is just this reason plus some additional consideration; more on this below.) If someone says that it's metaphysically impossible

for anything to exceed the speed of light, he either has to give us a compelling reason to think that it's metaphysically impossible for anything to exceed the speed of light, or show that there's something special about light that makes the suspension principle inapplicable to cases involving it. Since no one has ever done either, we're justified in refusing to think that faster-than-lightspeed travel is metaphysically impossible.

Likewise, if someone denies the metaphysical possibility of zombies, he either has to prove that the proposition that we have zombie twins is *necessarily* false, or show that there's something special about consciousness that makes the suspension principle inapplicable to cases involving it.

The most plausible attempt to show that it's necessarily false that we have zombie twins is a causation-based argument: (1) our conscious experiences have physical effects; (2) physical effects have only physical causes; so, (3) our conscious experiences are physical. Given the necessity of identity, it follows that it's metaphysically impossible for us to have zombie twins.[14]

The weakest link of this argument is (1).

Most non-philosophers do believe (1). If you ask them why, they'll say it's because whenever they're in pain (for example), their bodies act in certain ways. But the existence of such correlations doesn't establish any causal connection between pain and pain behavior; in particular, it doesn't show that pain causes pain behavior. Correlation doesn't imply causation. So, what most people take as evidence for (1) doesn't actually support (1) at all, any more than the correlation between cat populations

[14]See (Kirk, 1979) and (Papineau, 2002, 17-28).

and plague deaths in medieval European cities showed that cats caused plague.

Is there a better reason to accept (1)? Two reasons have been proposed.

First, there is the argument that we should accept that conscious experiences have physical effects, since that's the simplest explanation of the observed correlation between (e.g.) pain and pain behavior.[15]

Let's grant, for the sake of argument, that the simplest way to account for pain/pain-behavior correlations is by supposing that the pain causes the behavior. Well, there's also a correlation between plant growth and skin cancer: higher (or lower) rates of either come with higher (or lower) rates of both. The simplest explanation for this is that plant growth causes skin cancer (suppose we can exclude the reverse hypothesis, on the grounds that plants evolved before skin). But the simplest explanation is wrong. Plant growth doesn't cause skin cancer: sunshine causes both.

You might object that although the Cancer Plants hypothesis is the simplest explanation of the correlation between rates of plant growth and rates of skin cancer, it isn't part of the simplest explanation of *everything* under the Sun. That explanation, which takes into account all sorts of phenomena in addition to plant growth and skin cancer, explains the cancer and the verdure as effects of a common cause.

Fair enough, but why think it's any different when it comes to pain and pain-behavior? If you think that the simplest Theory of Everything will count pain behavior as an effect of pain, rather than as an effect

[15]Thus, for example, (Papineau, 2002, 23). The suggestion that it might always be the behavior that causes the pain rather than vice versa is ruled-out on empirical grounds.

(together with pain) of a common physical cause, that's only because you think that the Theory of Everything will count pain as a physical phenomenon. (If you didn't think that the Theory of Everything would count pain as a physical phenomenon, you wouldn't want to say that pain causes pain behavior, since that would commit you to overdetermination or a violation of the causal closure of the physical.) But whether we should count pain as a physical phenomenon is precisely the question that the causal argument was supposed to settle. The simplicity rationale for (1) is therefore question-begging.

A different argument in support of (1) is that if conscious experiences didn't have physical effects, we wouldn't know anything about conscious experiences, or even that there was such a thing as conscious experience. Since we obviously do know that we have conscious experiences, it follows that our experiences do have physical effects; or, so the argument goes.[16]

The most careful development of this argument is Robert Kirk's. According to Kirk, there are only three consideration-worthy explanations for how you can know about your own experiences (or at least, about their phenomenal qualities): (i) by your beliefs about the experiences being suitable effects of the experiences, (ii) by your beliefs about the experiences being suitable causes of the experiences, or, (iii) by your beliefs about the experiences being at least partially constituted by the experiences. Kirk argues that options (ii) and (iii) don't work, leaving us with (i) as the only viable account of our knowledge of consciousness.

[16]See (Watkins, 1989), (Kirk, 2005, 37-57), and the discussion in (Chalmers, 1996, 172-209).

It's not clear that an account along the lines of (ii) or (iii) couldn't be made to work, but let's grant that Kirk is right about that. It's also not entirely obvious that case (i) requires experiences to have physical effects; maybe we could make sense of the idea that our beliefs, or at least our justified true beliefs about our own experiences, are non-physical states of some sort. But let's set that aside too.[17]

The real problem with Kirk's argument, as I see it, is that it doesn't take into account a fourth way that one could know about one's own experiences.

If you think that a belief has to have an experience among its causes in order to count as knowledge of that experience, why is that? Presumably it's because you think that in order to count as knowledge, a belief that a certain experience occurred (or is occurring) must be a reliable sign or indicator that the experience occurred (or is occurring). After all, one's beliefs, including those that count as knowledge, have among their causes all kinds of things that are irrelevant to the beliefs' contents. If knowledge has to have its object among its causes, that's only because, or only to the extent that, such causation is necessary for reliable indication.

The point that Kirk overlooks is that it isn't necessary.

Take the standard epiphenomenalist picture, in which my belief that I'm in pain has no effects, but does have a cause—some brain event—that also causes my pain. In this case, the occurrence of my belief is as reliable an indicator of the occurrence of the pain as in the case in which the pain causes the belief. Or take my belief that I was in pain yesterday. If the

[17]Not setting it aside would mean getting sucked into the dreaded private language argument.

belief counts as knowledge, then, according to epiphenomenalism, it has among its causes a past brain-event that also caused a pain yesterday. In this scenario, the present occurrence of the belief is as reliable an indicator of the past occurrence of the pain as in a case in which the belief counts the pain itself among its causes.

I conclude that the causal argument against the metaphysical possibility of zombies fails, since we've been given no good reason to accept its first premise.

I said that there were two strategies by which one could try to block the suspension argument: by proving that zombies are metaphysically impossible, or by showing that our belief that zombies are metaphysically possible is the result of some kind of cognitive illusion. We've considered the first strategy and found it wanting. Now let's consider the second.

Advocates of the second strategy argue that there's something about our psychology that makes us peculiarly error-prone in our modal reasoning about consciousness, in a way that makes the suspension principle inapplicable to cases involving consciousness. I know of three attempts to show that we are prone to such error. They all take the form of debunking arguments: they're all arguments that try to cast doubt on our belief that zombies are metaphysically possible by showing that it arises from some factor that tends to make us have the belief regardless of whether zombies really are metaphysically possible.

The first debunking argument comes from Christopher Hill, building on a suggestion from Thomas Nagel.[18]

[18]For Nagel's suggestion, see (Nagel, 1974), and for the main developments thereof, (Hill, 1997) and (Hill & Mclaughlin, 1999).

Nagel distinguishes between two kinds of imagination: "perceptual" and "sympathetic."

> To imagine something perceptually, we put ourselves in a conscious state resembling the state we would be in if we perceived that thing. To imagine something sympathetically, we put ourselves in a conscious state resembling the thing itself.[19]

Nagel then suggests that we can debunk our intuition that a conscious mental state could occur without any corresponding physical state, by reference to the mutual independence of the two types of imagination:

> When we try to imagine a mental state occurring without its associated brain state, we first sympathetically imagine the occurrence of the mental state: that is, we put ourselves in a state that resembles it mentally. At the same time, we attempt to perceptually imagine the non-occurrence of the associated physical state, by putting ourselves into another state unconnected with the first: one resembling that which we would be in if we perceived the non-occurrence of the physical state. Where the imagination of physical features is perceptual and the imagination of mental features in sympathetic, it appears that we can imagine any experience occurring without its associated brain state and vice versa. The relation between them will appear contingent even if it is in fact necessary, because of the independence of the disparate types of imagination.[20]

A physicalist might try to account for our ability to form an undefeated conception of a zombie world in the same way, so as to discourage us from taking this ability as a reason to refrain from believing that a zombie world is metaphysically impossible.

---

[19](Nagel, 1974, 446).

[20]Ibid.

There are two questions, here. First: can we account for our ability to form an undefeated conception of a zombie world along the lines Nagel proposes? Second: if so, should that discourage us from taking our ability to form an undefeated conception of a zombie world as a reason not to believe that a zombie world is metaphysically impossible?

But we don't have to bother with the second question, since the answer to the first is "No." When I imagine a world that duplicates ours physically but contains no conscious experience, I don't do it by using sympathetic imagination. I just perceptually imagine a world with all the physical features of the actual world, and notice that nothing in this imaginative exercise required me to imagine any experience (perceptually, imaginatively, or otherwise). It's the same as when I perceptually imagine a car with all the standard features of my actual car, and notice that nothing in the imaginative exercise required me to imagine a sunroof (although it did require me to imagine an engine, a steering wheel, airbags, and various other things).

Hill suggests that when one conceives of a state of painlessness, one sympathetically imagines an absence of pain. Be that as it may, to conceive of a zombie world is to conceive of a world that contains no consciousness at all, painful or otherwise. But sympathetically imagining a total absence of experience is ruled-out by the very definition of sympathetic imagination. Anyway, even if one could use some kind of sympathetic imagination to conceive of a zombie world, one can also conceive of a zombie world using only perceptual imagination (or a combination of perceptual imagination and reflection upon perceptual imagination), as just explained.[21]

[21]For Hill's remarks on pain (or its absence), see (Hill, 1997, 69-70).

Another problem with Hill's argument is that it simply fails to persuade. That's an unfortunate feature for any argument to have, but for a debunking argument, it's fatal.

Suppose you make it clear to me that the only reason I believe that my children are above average is that I tend to exaggerate evidence that they're superior to other children and discount evidence to the contrary, and suppose you also make it clear to me that I have this tendency independent of my children's actual merits. Then, as a rational and clear-thinking person, I'll stop believing that my children are above average; or, if for some reason I can't get rid of the belief, I'll at least be very uncomfortable with it, and do what I can to suspend or disavow it. By the same token, if you tell me something that leaves my belief in the superiority of my children intact and undisavowed, it follows (assuming that I'm rational and clear-thinking) that you haven't successfully debunked my belief—haven't shown that it arises from a factor that's insensitive to my children's actual merits (or lack thereof).

The point is a general one. Suppose that a rational and intelligent person, N, believes that p, but that his belief that p is attributable to some factor, X, that's independent of whether it's true that p (so, N's belief that p doesn't depend on its being true that p). If you now reveal to N that this is the situation—i.e., that his belief that p is attributable to something that gives rise to the belief regardless of whether it's true that p—N will stop believing that p, or, if for some reason he can't stop believing that p, he'll come to regard his belief that p as an alien and unwelcome presence in his mind, like a persistent but groundless suspicion that he's being surveilled by the FBI.[22]

[22]Andrew Melnyk suggests that our intuition that zombies are possible might be like the Müller-Lyer illusion, in which two lines that we know are equally long appear unequal. (Melnyk, 2002) But in the Müller-Lyer case, I'm not at all inclined

I doubt that the Nagelian debunking argument has had this effect on anyone, physicalist or antiphysicalist, with respect to the belief that zombies are metaphysically possible. If it has, then why do rational and intelligent people keep on arguing about zombies? Anyway, the Nagelian argument definitely hasn't had this effect on me, from which it follows that my use of two kinds of imagination in my thinking about consciousness (perceptual and sympathetic) doesn't account for my belief that zombies are metaphysically possible, or at least not in a way that casts doubt on that belief.[23]

I don't think I'm setting the bar too high here. It is possible to debunk a modal intuition. Kripke did it with the intuition that water could have been something different from H_2O. All I'm saying is that if a physicalist wants to debunk our intuition that zombies are metaphysically possible, he needs to follow Kripke's example.

A different debunking argument also inspired by Nagel's comments comes from David Papineau. Papineau argues that we believe that zombies are metaphysically possible only because we commit a special kind of use-mention fallacy, which he calls the "antipathetic fallacy." Again, the idea is that we have two ways of thinking about conscious experience. Sometimes we think about conscious experience partly by *using* conscious experience, and sometimes we think about conscious experience without using it. (Analogously, I can think about a chess position partly

---

to *believe* that the lines are unequal; it's not like I have to keep reminding myself that they're really the same length, on pain of slipping into the belief that they're unequal. Once I put a ruler to the things, I lose any inclination to believe that they differ in length. By contrast, noting that I employ different modes of imagination in my thinking about consciousness—putting a ruler to my imaginative faculties, so to speak—does nothing to diminish my inclination to believe that zombies are possible.

[23] In Stephen Yablo's terms, the Nagelian debunking argument fails to satisfy "the psychoanalytic standard"; see (Yablo, 2008, 159).

by using a chessboard set up in that position, or I can think of the position without using a chessboard; e.g., by thinking of it in terms of the standard algebraic notation.) Because we sometimes use consciousness to think about consciousness, we are apt to think that we *aren't* thinking about consciousness when we don't use consciousness to think about it, when really we might just be thinking about consciousness in a way that doesn't involve using consciousness—in a way that mentions pain, for example, without using any pain-related phenomenology. Or, so Papineau argues.[24]

Papineau's argument is subject to the same objections as Hill's.

Suppose we grant that we often use ϕ (or ϕ-related) experience to conceive of experience with ϕ phenomenology. How does that shed light on what happens when we conceive of a total absence of experience, as we do when conceiving of zombies? In this case, there's no relevant phenomenology for us to use to conceive of the contemplated state of affairs.

In any event, Papineau's argument fails to shake our belief in the metaphysical possibility of zombies, with the fatal results already mentioned in connection with Hill's argument. If I believe that p only because I commit some use-mention fallacy, and you point this out to me, I'll stop believing that p. This is particularly so, considering that use-mention fallacies are generally hard to fall for and easy to avoid once pointed out. Since I'm still comfortably believing that zombies are metaphysically possible,

[24]See (Papineau, 2002, 161-74). This style of debunking argument is pretty popular, with versions of it espoused by (Loar, 1990, 90), (Tye, 1999), and (Perry, 2001, 119-50), among others. I focus on Papineau, since his version of the argument is the most detailed.

I have to conclude that Papineau has failed to show that I'm believing it only because I'm committing a use-mention fallacy.[25]

The third debunking argument to consider comes from Andrew Melnyk, who picks up on a suggestion discarded by Papineau. The suggestion is that believing a (non-trivial) identity claim requires combining two mental representations in a way that involves something like "mental file merging," where such merging is psychologically impossible, when one of the representations is phenomenal, and the other non-phenomenal.[26]

Why are phenomenal representations supposed to be psychologically unmergeable with non-phenomenal (material) ones?

> The reason I have in mind is that one kind of phenomenal concept seems to be usable only to refer to a phenomenal state as one undergoes it ("*That* is going on in me now"), and not to be usable to *re-identify* a phenomenal state, not even to re-identify it as *one of those again*. Now if phenomenal concepts of this kind exist, and if concepts in general can be viewed as analogous to files, then a phenomenal concept of this kind will constitute a file that is only temporary, a file that persists only as long as one is undergoing the experience it picks out. But any file corresponding to a material concept will presumably be permanent; at the very least it will per-

[25]The chess analogy points to another problem. If Papineau's explanation of my intuition that zombies are metaphysically possible is correct, why don't I also intuit that it's possible for a chess board to be set up in a certain position in a world in which nothing satisfies the corresponding algebraic description? Pär Sundström argues that Papineau's account even predicts intuitions of *phenomenal* distinctness where we have none; see (Sundström, 2008, 141-42).

[26](Melnyk, 2002), citing (Papineau, 2002, 165). A phenomenal representation of an experience—or "phenomenal concept," in Papineau's terminology—is one that uses that experience, or phenomenology related to it, to represent the experience. Non-phenomenal concepts are what Papineau calls "material concepts."

> mit the re-identification of whatever it picks out. And, on the not too implausible assumption that no temporary file can be merged with a permanent file, it follows that no phenomenal concept of the kind in question can be merged with a material concept, and hence, if believing identity claims is a matter of mental file-merging, that no identity claim framed using a phenomenal concept of the kind in question and a material concept can be believed.[27]

One question this raises is why a temporary file couldn't merge with a permanent one. Why can't I use a material concept of pain to think about pain while I'm in pain, and also thinking about it as *this* unpleasant experience? (As in, "*this* pain is accompanying such-and-such brain activity.") Maybe the phenomenal representation of pain isn't as enduring as the material representation, but so what? I can merge two computer files, even if one of them but not the other is infected with a virus that will soon delete the file. I can merge two paper files, even if one of them is printed on cheap newsprint and the other on vellum.

The more serious problems with Melnyk's proposal are the same as those that pertained to the other debunking arguments we've considered. Even if believing a statement of the form, "*This* sort of experience = such-and-such neural process" would require a psychologically impossible mental file merger, how does that explain our ability to believe a statement such as: "There could be a non-conscious being physically just like me"? And—most important of all—if Melnyk's explanation of why zombies seem metaphysically possible to us is correct, why don't we lose our inclination to believe that they're possible when we reflect on Melnyk's proposal?

---

[27] (Melnyk, 2002).

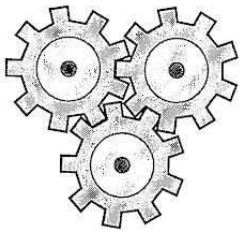

Figure 1: Unworkable gearing

Suppose I draw plans for a complicated machine containing lots of gears, cams, levers, push-rods, etc. The best way to establish the design's workability would be to build an actual machine according to the plans, but maybe that's prohibitively expensive or beyond the capabilities of existing manufacturing technology (like Charles Babbage's Analytical Engine). However, I've taken great care in drawing the plans, and after studying them carefully, the design seems workable—I don't see anything mechanically impossible about it (like the gear assembly in Fig. 1). I show the design to my engineer friends and they see no mechanical impossibility in it either. Finally, I consult a panel of clinical psychologists, asking them whether they can think of any perceptual or cognitive illusion that might have blinded me or the engineers to a mechanical impossibility lurking in my design. They know of none. (One of the psychologists notes that a moiré pattern can give rise to optical illusions, but I point out to him that my plans contain no moiré pattern.)

Given all this, am I entitled to judge that my design is mechanically possible? That's debatable, perhaps; but I'm certainly *not* entitled to think that the design is mechanically *impossible*. Rather, I'm entitled to find fault with anyone who asserts that the design is mechanically impossible without producing some heretofore unknown proof of its mecha-

nical impossibility, or pointing out some heretofore unnoticed optical illusion in my plans.

I submit that the situation is exactly the same when it comes to the metaphysical possibility of zombies. If you want to say that the physical facts of our world metaphysically entail the mental facts about it—for example, if you want to advocate physicalism—you better be prepared to back your statement up with some heretofore unknown demonstration that a zombie world is metaphysically impossible, or a debunking argument that actually makes people stop believing that zombies are possible (or at least makes them try to stop believing it). If you can't do either of those things, you shouldn't be a physicalist.

5 Why we should believe that physicalism is false

The suspension argument revises the naive untenability argument by weakening its conclusion. Is there a way to revise the naive untenability argument that preserves its conclusion?

I think so.

First, notice that when it comes to the question whether something could, as a matter of metaphysical possibility, travel faster than light, we don't merely suspend judgement; we positively judge that faster-than-light travel is metaphysically possible. This judgement isn't some weird cognitive tick, either. We consider ourselves *justified* in having the belief. Why?

Earlier, we considered the following answer: because we can conceive of something traveling faster than light, and know of no proof that it's metaphysically impossible for something to travel faster than light. The problem with this answer was that when applied to other cases (like the Goldbach conjecture), it wrongly implied that we'd be justified in accep-

ting claims that we know (or, should know) are mutually incompatible. Still, *something* has got to justify our belief that superluminal travel is metaphysically possible, since the belief clearly is justified.

The key question here is: what relevantly distinguishes the proposition that the Goldbach conjecture is true from the proposition that something travels faster than light?

The obvious answer is that as far as we know, the proposition that something travels faster than light, unlike the proposition that the Goldbach conjecture is true, is neither necessarily true nor necessarily false (metaphysically necessarily, that is). In this respect, the proposition about light also differs from the proposition that the Goldbach conjecture is false, the proposition that water is HO, the proposition that water is H_2O, the proposition that Hesperus is Phosphorus, and the proposition that Hesperus isn't Phosphorus. Unlike these other propositions, we have no good reason to think that the proposition that something travels faster than light is non-contingent. (A proposition is non-contingent just in case it is necessarily true or necessarily false.)

So, I suggest that the "something" that justifies our belief that superluminal travel is metaphysically possible is a combination of two factors: (1) our ability to form an undefeated conception of superluminal travel, and, (2) our lack of any good reason to think that the proposition that something exceeds the speed of light is non-contingent.

The general principle here is: if (a) you can conceive of its being the case that p, (b) you know of no proof that it's metaphysically impossible that p, and (c) you have no good reason to think that the proposition that p is non-contingent, then you should believe that it's metaphysically possible that p. Call this the *untenability principle*.

We have good reasons to think that "The Goldbach conjecture is true," "The Goldbach conjecture is false," "water is HO," "water is H_2O," "Hesperus is Phosphorus," "Hesperus isn't Phosphorus" are non-contingent. So the untenability principle doesn't tell us that we should believe that it's possible that the Goldbach conjecture is true, or possible that the Goldbach conjecture is false, or possible that water is HO, or possible that water is H_2O, etc., even if we can conceive of these, and even if we know of no proof that they are metaphysically impossible.

But the principle does tell us that we should believe that it's possible for something to travel faster than light, since here clauses (a) through (c) are all satisfied. In particular, the proposition that something travels faster than light satisfies (c), since we have no good reason to think that this proposition is (metaphysically) necessarily true or (metaphysically) necessarily false.

What about zombies? We can form an undefeated conception of them. But do we have a good reason to think that the proposition that we have zombie twins is non-contingent?

Obviously, we have no good reason to think that this proposition is necessarily true, since we have no good reason to believe that it's true at all—presumably we don't actually have physical duplicates, conscious or non-conscious. So the question comes down to this: do we have any good reason to think that it's necessarily false that we have zombie twins? If so, the untenability principle doesn't tell us that we should believe that zombies are possible. If not, it does.

At this point, there's a danger of getting bogged down in a lengthy discussion about what counts as a "good reason" for believing that a proposition is non-contingent (e.g., necessarily false). But all we really need to do here is review the considerations that physicalists offer as reasons

for thinking that it's necessarily false that we have zombie twins, and point out that when it comes to ostensible metaphysical possibilities that pose no threat to physicalism, the physicalist wouldn't accept parallel considerations as good reasons for denying that those are genuine metaphysical possibilities. (Of course, a physicalist can always suggest that there's something about consciousness that makes such parallels irrelevant, but we've already considered the best attempts to make good on that suggestion, and found them wanting.)

One good reason for believing that it's necessarily false that there are zombies would be a proof that zombies are metaphysically impossible. We've already considered and rejected the best attempt at such a proof (the causal argument). It remains to consider less-than-probative reasons, or ostensible reasons, for believing that zombies are metaphysically impossible.

There's really just one: gains in simplicity. As Ned Block and Robert Stalnaker point out, a major reason for making an empirical identification is that doing so makes for better explanations of natural phenomena and an overall simpler world-view:

> Suppose that heat = molecular kinetic energy, pressure = molecular momentum transfer and boiling = a certain kind of molecular motion . . . Then we have an account of how heating water produces boiling. If we were to accept mere correlations instead of identities, we would only have an account of how something correlated with heating causes something correlated with boiling. Further, we may wish to know how it is that increasing the molecular kinetic energy of a packet of water causes boiling. Identities allow a transfer of explanatory and causal force not allowed by mere correlations. Assuming that heat = mke, that pressure = molecular

> momentum transfer, etc., allows us to explain facts that we could not otherwise explain. . .
>
> If we believe that heat is correlated with but not identical to molecular kinetic energy, we should regard as legitimate the question of why the correlation exists and what its mechanism is. But once we realize that heat *is* molecular kinetic energy, questions like this will be seen as wrongheaded.[28]

Likewise, if we identify conscious experiences with neural processes, we get a simpler account of the phenomena, and quash demands for an explanation of psychophysical correlations.

Physicalists promoting their view on grounds of simplicity often write as though science always gives simplicity more weight than "mere" prima facie metaphysical possibility. It seems to me that this misrepresents actual scientific thinking.

Our universe began (circa the Big Bang) in a state of extremely low entropy. Statistically, the likelihood of matter being distributed in such a low-entropy way is mind-bogglingly small (Roger Penrose estimates it at 1 in $10^{10^{123}}$). Yet, as far as we know, a different distribution of matter wouldn't have been nomologically impossible, let alone logically or metaphysically impossible. Everyone agrees, or should agree, that there are metaphysically possible worlds whose initial states aren't low-entropy states. This is why it seems reasonable to ask for an explanation for why the initial state of *our* world had such very low entropy.[29]

---

[28](Block & Stalnaker, 1999, 23-24).

[29]Whether it's reasonable to expect an answer better than "that's just the way it is" is another question. For good philosophical discussion of the low-entropy past, see (Price, 1996), (Price, 2004), and (Callender, 2004). For Penrose's estimate, see (Penrose, 1989, 344).

Of course, it would simplify things if we equated "being a very low entropy state" with "being the initial state of a physical universe." Then the question, "Why was entropy so low circa the Big Bang?" would simply not arise. It would be like asking why heat correlates with molecular energy.

Yet, this doesn't incline us to think that it's metaphysically impossible for a physical universe to begin in anything but a low-entropy state. So why should the simplification we'd get by equating consciousness with brain processes incline us to think that it's metaphysically impossible for a physical copy of me to lack consciousness?[30]

There's a mystery about the low-entropy past, and there's a mystery about consciousness. If high-entropy initial states were metaphysically impossible, the low-entropy past would have a simple explanation, and the first mystery would dissolve. If zombies were metaphysically impossible, the existence of consciousness would have a simple explanation, and the second mystery would dissolve. But we don't think that the simplifying and mystery-dissolving benefits of denying the metaphysical possibility of high-entropy initial states entitle us to deny the metaphysical possibility of high-entropy initial states. So we shouldn't think that the simplifying and mystery-dissolving benefits of denying the metaphysical possibility of zombies entitle us to deny the metaphysical possibility of zombies. Ockham's Razor just doesn't cut it here.

[30]One might suggest that we *should* equate being the initial state of a universe with being a state of low entropy, on the grounds that we should define temporal order in terms of a monotonic entropy gradient. This is objectionable for various reasons, not least of which is that it implies that there's no temporal order in a universe bookended by low-entropy states. The important point for present purposes is that we don't reject the metaphysical possibility of such a universe just because that would allow us to accept an entropic analysis of temporal order.

I've argued that we can conceive of a zombie world, that we know of no proof that a zombie world is metaphysically impossible, and that we have no good reason to think that the proposition that there are zombies is non-contingent. This yields the following argument: (1) we can conceive of a zombie world, and we know of no proof that such a world is metaphysically impossible, and we have no good reason to think that the proposition that there are zombies is non-contingent; (2) if we can conceive of a zombie world, know of no proof that such a world is metaphysically impossible, and have no good reason to think that the proposition that there are zombies is non-contingent, then we should believe that a zombie world is metaphysically possible; therefore, (3) we should believe that a zombie world is metaphysically possible. Since the third conjunct of (1) entails the second, we can state the argument more succinctly as follows:

E1 We can conceive of a zombie world, and we have no good reason to think that the proposition that there are zombies is non-contingent.

E2 If we can conceive of a zombie world, and we have no good reason to think that the proposition that there are zombies is non-contingent, then we should believe that zombies are metaphysically possible.

E3 So, we should believe that zombies are metaphysically possible.

Call this the *refined untenability argument*. Unlike the naive untenability argument, it can't be spoofed with arguments purporting to show that we should believe that the Goldbach conjecture is possibly true and possibly false. Unlike the suspension argument, it shows not merely that we shouldn't be physicalists, but that we should be anti-physicalists.

Like the suspension argument, this one would collapse in the face of a proof that zombies are metaphysically impossible, or a successful debunking of the intuition that zombies are metaphysically possible. Until then, I submit that we have as much reason to accept the refined untenability argument as we have to believe in the metaphysical possibility of any uncontroversial non-actual state of affairs.

6 Kripkean arguments

The refined untenability argument doesn't apply directly to the identity theory, since we do have reason to think that "pain is not C-fiber stimulation" is non-contingent (at least, if we take the standard view that this is a statement of a non-identity, and that such statements, like identity statements, are non-contingent). But the argument does apply to the identity theory indirectly, since the identity theory entails that it's metaphysically impossible for there to be a physical duplicate of our world that contains no conscious experience, whereas the refined untenability argument tells us we should believe that such a world is metaphysically possible.

There is an argument that addresses the identity theory directly, without, however, trying to show that the identity theory is false, or that you should believe that it's false, or even that you shouldn't believe that it's true. It's an argument for the claim that we do not, in fact, believe that the identity theory is true.

The argument I have in mind is close to one that David Papineau reconstructs from Saul Kripke's well-known discussion of the identity theory. According to Papineau,

> Kripke's challenge isn't to explain how mind-brain identities are a posteriori—as it were, to explain how they can appear possibly false to people who don't yet believe them. Rather his challenge is

to explain why they *still* appear possibly false, even to people who *do* believe them.[31]

In our discussion of the refined modal argument against physicalism (in §2), we raised the point, due to Kripke, that once it was settled that water was H_2O, erstwhile proponents of the water-is-HO hypothesis would have realized that what they formerly took to be conceiving of water as HO was really just conceiving of a situation in which something that outwardly resembled water is HO. Similarly, once you become convinced that Hesperus is Phosphorus (or Cicero Tully, or whatever), you no longer consider yourself able to conceive of a situation in which Hesperus isn't Phosphorus. Now that you're convinced that Phosphorus is the same thing as Hesperus, you simply have no way to conceive of a situation in which Phosphorus isn't the same thing as Hesperus, since that would require you to conceive of a situation in which a thing was different from itself.

By the same token, if you were convinced that one of your pains was the same thing as some brain state (say, a stimulation of your C-fibers), then you'd have no way to conceive of a situation in which the pain wasn't C-fiber stimulation (e.g., a zombie world). You would look upon anything that you might once have considered conceiving of such a situation as really just a case in which something outwardly resembling pain isn't C-fiber stimulation. However, as Kripke notes, anything that outwardly resembles pain *is* pain.

The upshot is that *if you think you can conceive of a zombie world, then you don't believe that consciousness is a brain process*. This yields the following argument, which I'll call the *omissive Kripkean argument*:

[31](Papineau, 2007, 479). For an earlier construal of Kripke's challenge along the same lines, see (Levine, 1983).

F1 If you believe that consciousness is brain process, then you don't think you can conceive of a zombie world.

F2 You do think you can conceive of a zombie world.

F3 So, you don't believe that consciousness is a brain process.

As I've presented it, the Kripkean argument concludes that you don't believe that consciousness is a brain process. In Papineau's version of the argument, the conclusion is that you *do* believe that consciousness is *not* a brain process; let's call this the *commissive Kripkean argument*.[32]

The commissive argument replaces F1 with the premise that if you think you can conceive of a zombie world, then you believe that consciousness isn't a brain process. But you might think that you can conceive of a zombie world, yet refrain from believing that consciousness isn't a brain process, because you know that there are cases in which people have taken themselves to conceive of things that turned out to be metaphysically impossible. Still, if you think you can conceive of a zombie world, and you possess a modicum of rationality, you won't believe that consciousness *is* a brain process, since in order to believe that, you'd have to think that you were conceiving of a thing as differing from itself. So, even if the commissive version of Kripke's argument is open to doubt, the omissive version stands.

Papineau, himself a physicalist, accepts the commissive argument, but points out that its conclusion is consistent with the claim that some us believe that zombies are impossible. Physicalists just have inconsistent beliefs: they believe that zombies are possible, and they believe that zombies aren't possible. But (according to Papineau) that makes the situation sound worse than it necessarily is. At a theoretical level—when

[32]See (Papineau, 2007, 478-80).

thinking about the scientific advantages of physicalism—we're inclined to deny that zombies are possible, while at an intuitive level—when not thinking about the scientific advantages of physicalism—we're inclined to believe that zombies are possible. The question is which level is more important. A physicalist will presumably say it's the theoretical level.

If this response works for the commissive argument, it works for the omissive argument too. On a theoretical level, we believe that zombies are impossible; on an intuitive level, we don't. If the theoretical level is more important, a physicalist might learn to live with this tension.

But is the theoretical level more important than the intuitive level? Not always (recall the low-entropy past). So why think it's more important in this particular case? If you say it's because modal intuitions about consciousness are unreliable, you owe us a good debunking argument.

There's another way a physicalist might try to accommodate the commissive argument.

Belief isn't, or needn't be, all-or-nothing; it comes in degrees. If we look at the situation in terms of graded beliefs ("credences") instead of all-or-nothing ("binary") beliefs, we can say, on the physicalist's behalf, that the commissive argument shows only that physicalists don't assign as high a credence to physicalism as they might like to. That doesn't mean that they don't assign higher credence to the proposition that physicalism is true than to the proposition that it's false.

An immediate problem with this credence-splitting strategy is that it doesn't apply to the omissive version of Kripke's argument, which concludes not that you *have* the belief that zombies are possible, but that you *lack* the belief that zombies *aren't* possible. Here, there's no question of dividing your credences between two propositions—that zombies are

possible, and that they aren't—since (if the argument is sound) you don't believe the second proposition at all.

A physicalist might say that the conclusion of the omissive argument should really just be that we don't have a very high (greater than 0.8?) credence that consciousness is a brain process. That is, he might insist on replacing F1 with the premise that if we think we can conceive of a zombie world, then we don't believe *to a very high degree* that consciousness is a brain process. That would leave open the possibility that we believe to a degree greater than 0.5 that consciousness is a brain process.

Suppose that, like most people who have an opinion about such matters, you think you can conceive of a situation in which William Shakespeare exists but Christopher Marlowe doesn't. You've considered the possibility that all you can really conceive of is a situation in which someone who outwardly resembles Shakespeare exists in a Marloweless world, but you are confident that that's not what you're doing. You've reviewed all the evidence that people have offered in support of the theory that Shakespeare and Marlowe were actually the same man (e.g., putative evidence that the works commonly attributed to Shakespeare were actually penned by Marlowe), and found that none of it holds up to scrutiny.

Where does that put your credence that Shakespeare was Marlowe? Above 0.5? Presumably not. But then presumably your credence that consciousness is a brain process isn't above 0.5 either, if you think you can conceive of a situation in which the brain process exists but consciousness doesn't.

Conclusion

We've considered a variety of broadly anti-physicalist arguments, including arguments that purport to show that physicalism is false, arguments

that purport to show that we should believe that physicalism is false, arguments that purport to show that we should not believe that physicalism is true, and arguments that purport to show that we do not, in fact, believe that physicalism is true.

The good news for the physicalist is that arguments of the first, most aggressive type don't pose a clearly lethal threat to his position. The bad news is that each of the other three types of argument exists in at least one version that the physicalist has no evident way of blocking. All of these arguments—the suspension argument, the refined untenability argument, and the omissive version of Kripke's argument—are hostage to fortune, in that their success depends on nobody's proving that zombies aren't possible, or successfully debunking our intuition that they are. I've argued that the best attempts at such a proof or debunking have failed. Until and unless someone comes up with a better response to these arguments, the balance of considerations inclines steeply in favor of rejecting physicalism, or at the very least declining to embrace it.

References

Block, Ned, & Stalnaker, Robert. 1999. Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review*, **108**(1), 1–46.

Blumson, Ben, & Tang, Weng Hong. 2015. A Note on the Definition of Physicalism. *Thought: A Journal of Philosophy*, **4**(1), 10–18.

Callender, Craig. 2004. There is no puzzle about the low-entropy past. *Pages 240–255 of:* Hitchcock, Christopher (ed), *Contemporary Debates in Philosophy of Science*. Malden: Blackwell.

Chalmers, David. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Chalmers, David, & Jackson, Frank. 2001. Conceptual Analysis and Reductive Explanation. *Philosophical Review*, **110**(3), 315–361.

Chalmers, David J. 2002. Does conceivability entail possibility? *Pages 145–200 of:* Gendler, Tamar Szabó, & Hawthorne, John (eds), *Conceivability and Possibility*. Oxford: Clarendon Press.

Goff, Philip. 2017. *Consciousness and Fundamental Reality*. New York: Oxford University Press.

Hill, Christopher. 1997. Imaginability, conceivability, possibility and the mind-body problem. *Philosophical Studies*, **87**(1), 61–85.

Hill, Christopher, & Mclaughlin, Brian. 1999. There are fewer things in reality than are dreamt of in Chalmers's philosophy. *Philosophy and Phenomenological Research*, **59**(2), 445–454.

Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Clarendon Press.

Kirk, Robert. 1979. From physical explicability to full-blooded materialism. *Philosophical Quarterly*, **29**(116), 229–237.

Kirk, Robert. 2005. *Zombies and Consciousness*. Oxford: Clarendon Press.

Kripke, Saul. 1980. *Naming and Necessity*. Oxford: Basil Blackwell.

Kung, Peter. 2010. Imagining as a guide to possibility. *Philosophy and Phenomenological Research*, **81**(3), 620–633.

Levine, Joseph. 1983. Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, **64**(4), 354–361.

Loar, Brian. 1990. Phenomenal states. *Philosophical Perspectives*, **4**, 81–108.

Melnyk, Andrew. 2002. Papineau on the intuition of distinctness. *SWIF Philosophy of Mind*, **4**(1).

Nagel, Thomas. 1974. What is it like to be a bat? *The Philosophical Review*, **83**(4), 435–450.

Papineau, David. 2002. *Thinking about Consciousness*. Oxford: Clarendon Press.

Papineau, David. 2007. Kripke's proof is ad hominem not two-dimensional. *Philosophical Perspectives*, **21**, 475–494.

Penrose, Roger. 1989. *The Emperor's New Mind*. Oxford: Oxford University Press.

Perry, John. 2001. *Knowledge, Possibility, and Consciousness*. Cambridge: MIT Press.

Price, Huw. 1996. *Time's Arrow and Archimedes' Point: New Directions for the Physics of Time*. Oxford: Oxford University Press.

Price, Huw. 2004. On the origins of the Arrow of Time: why there is still a puzzle about the low-entropy past. *Pages 219–239 of:* Hitchcock, Christopher (ed), *Contemporary Debates in Philosophy of Science*. Malden: Blackwell.

Stoljar, Daniel. 2006. *Ignorance and Imagination: The Epistemic Origin of the Problem of Consciousness*. New York: Oxford University Press.

Stoljar, Daniel. 2010. *Physicalism*. New York: Routledge.

Sundström, Pär. 2008. Is the mystery an illusion? Papineau on the problem of consciousness. *Synthese*, **163**(2), 133–143.

Tye, Michael. 1999. Phenomenal consciousness: the explanatory gap as a cognitive illusion. *Mind*, **108**(432), 705–725.

Watkins, Michael. 1989. The knowledge argument against 'the knowledge argument'. *Analysis*, **49**(3), 158–160.

Yablo, Stephen. 2008. *Thoughts: Papers on Mind, Meaning, and Modality*. Oxford: Oxford University Press.